GCNs-Based Context-Aware Short Text Similarity Model
Xiaoqi Sun, Shaochun Wu, Yue Liu
Shanghai University

Abstract

Semantically textual similarity is a fundamental task in text mining and natural language processing (NLP), which has profound research value. The essential step for text similarity is text representation learning. Recently, researchers have exploited the graph convolutional network (GCN) techniques on text representation, since GCN does well in handling complex structures and presenting sythetic information. However, current GCN models are usually limited to very shallow layers due to the vanishing gradient problem, which can not capture non-local dependency information of sentences. In this paper, we propose a GCNs-based context-aware (GCSTS) model that applies iterated GCN blocks to train deeper GCNs. Recently employing the same GCN block prevents over-fitting and provides broad effective input width. Combined with dense connections, GCSTS can be trained more deeply. Besides, we use dynamic graph structures in the block, which further extends the receptive field of each vertex in graphs, learning better sentence representations. Experiments show that our model outperforms existing models on several text similarity datasets, while also verify that GCNs-based text representation models can be trained in a deeper manner, rather than being trained in two or three layers.

Introduction

• Recently, Graph Convolutional Network (GCN) has attracted extensive attention. GCNs can extract semantic and syntactic information of sentences represented from sentence dependency trees.
• GCNs capture information only about immediate neighbors with one layer of convolution. L layers will be needed in order to capture neighborhood information that is far away. Therefore, a shallow GCN model can not be able to capture non-local interactions of long sentences.
• Normally, when dealing with a large graph, the neighborhood information of a graph, current GCN models used in NLP are no more than three layers. Due to the fact that the deeper GCN model introduces a higher complexity in backpropagation and the vanishing gradients pose limitations on the depth growth of GCNs-based networks.
• Furthermore, most GCNs employ fixed graph structures. However, dynamic graph convolution, where the graph structure is allowed to change in each layer, can extend the receptive field of each vertex in the graph and learn a better graph representation compared to GCNs with fixed graph structure.
• On the basis of analysis above, we propose a deeper GCNs-based context-aware network to get better text representations. Our overall iterated GCN architecture reaps the same block of graph convolution for various sentence representations. Besides, we use dynamic graph structures in the GCN block. The dynamic graph convolution extends the representations of nodes influenced by non-fixed neighborhoods, which means the representations can bring more contextual information.
• In order to alleviate the vanishing gradient problem in deeper GCNs, we adopt a similar idea of DenseNet

Results

Table 1 STS12-STS16: SemEval 2012-2016 datasets. STS-B: STSbenchmark dataset. We report the Spearman correlation coefficient in this work. The best results are bold.

Table 2 Experimental results on the MRPC dataset. The best results are bold.

Table 3 An ablation study for the proposed model.

Conclusion

• In this paper, we propose a GCNs-based context-aware short text similarity model. GCSTS uses iterated GCN blocks and non-fixed neighbor matrices to incorporate more contextual information of sentences. The experimental results show that the proposed model can effectively improve the performance of sentence similarity measurement and provides a new way to train deeper GCNs.

Acknowledgement

• This work is supported by the State Key Program of National Science Foundation of China (No. 61936009) and the National Natural Science Foundation of China (No.52073169). We also appreciate the High Performance Computing Center of Shanghai University and Shanghai Engineering Research Center of Intelligent Computing System (No. 19DZ2252600) for providing the computing resources and technical support.